

## **WHAT IS CLAIMED IS:**

1. A method for elucidating a protein expression profile of a test cell line or group of cells, the method comprising:

randomly introducing into the genome of a cell or group of cells a promoterless polynucleotide construct, the construct comprising in a 5' to 3' orientation:

- i) a splice acceptor consensus sequence;
- ii) a complementary sequence of a first type IIS restriction enzyme recognition sequence;
- iii) an oligonucleotide sequence encoding an assayable marker peptide;
- iv) a sequence of a second type IIS restriction enzyme recognition sequence;
- v) a splice donor consensus sequence;

wherein said promoterless polynucleotide construct when introduced into an actively expressed gene results in the generation of a fusion protein, containing the assayable marker peptide inserted at a random position within two exons coding for the cellular protein encoded by said gene;

- vi) identifying those cells expressing said marker peptide fused to said cellular protein;
- vii) determining the identity of the proteins to which the marker peptide is fused in each group of cells.

2. A method to identify differentially expressed proteins in two different populations of cells, the method comprising:

randomly introducing into the genomes of a reference group of cells and into the genomes of a test group of cells a promoterless polynucleotide construct, wherein the construct comprises, in a 5' to 3' orientation;

- i) a splice acceptor consensus sequence;
- ii) a complementary sequence of a first type IIS restriction enzyme recognition sequence;

- iii) an oligonucleotide sequence encoding an assayable marker peptide;
- iv) a sequence of a second type IIS restriction enzyme recognition sequence;
- v) a splice donor consensus sequence;

thereby generating a population of cells, each bearing a different cellular protein with an assayable peptide marker inserted at random positions within each fusion protein;

- vi) identifying those cells expressing said marker peptide fused to said cellular protein in each group of cells;
- vii) determining the identity of the proteins to which the marker peptide is fused in each group of cells; and
- viii) comparing by statistical methods the protein expression profiles obtained for the test group of cells against the protein expression profiles obtained for the reference group of cells, thereby identifying differences in the expression levels of fusion proteins among the two groups of cells.

3. The method of claim 1 further comprising sorting cells identified in step vi) into monoclonal subgroups of genetically identical cells.

4. The method of claim 1 further comprising sorting cells identified in step vi) into oligoclonal subgroups of cells based on their differential levels of expression of the marker peptide.

5. The method of claim 2 further comprising sorting cells identified in step vi) into monoclonal subgroups of genetically identical cells.

6. The method of claim 2 further comprising sorting cells identified in step vi) into oligoclonal subgroups of cells based on their differential levels of expression of the marker peptide.

7. The method of claim 1, 2, 3, 4, 5 or 6 wherein the identity of the protein fused to the marker peptide is determined using a method selected from the group consisting of 5' Rapid Amplification of cDNA Ends (5' RACE), 5' Serial Analysis of Viral Integration (5'SAVI), 3' Serial Analysis of Viral Integration (3'SAVI) and Serial Analysis of Viral Integration (SAVI).

8. The method of claim 1, 2, 3, 4, 5 or 6 wherein the identity of the protein fused to the marker peptide is determined using a method of Serial Analysis of Viral Integration (SAVI), wherein the said method consists of:

- i) isolating mRNA from each subgroup of cells;
- ii) reverse transcribing the mRNA into double stranded cDNA;
- iii) subjecting the cDNA to a reaction with at least one enzyme that recognizes the type IIS restriction enzyme recognition sequences flanking the oligonucleotide encoding the peptide marker sequence, and cleaves the cDNA upstream of the first recognition sequence and downstream of the second recognition sequence, thereby generating one or more cDNA fragments, wherein each of these fragments comprise the oligonucleotide sequence corresponding to the upstream and downstream exons directly fused to the marker peptide encoding sequence;
- iv) self-ligating all cDNA fragments generated by digestion with said restriction enzyme;
- v) amplifying by the inverse polymerase chain reaction the fragments containing the oligonucleotide sequences of the exons fused to the marker peptide;
- vi) cloning and sequencing said amplified fragments; and
- vii) comparing the sequence of each oligonucleotide against oligonucleotide sequences in a one or more nucleotide sequence database thereby identifying one or more fusion proteins present in each subgroup of cells.

9. The method of claim 1, 2, 3, 4, 5 or 6 wherein the identity of the protein fused to the marker peptide is determined using a method of 5'Serial Analysis of Viral Integration (5'SAVI), wherein said method consists of:

- i) isolating RNA from each subgroup of cells;
- ii) reverse transcribing the mRNA into double stranded cDNA;
- iii) subjecting the cDNA to a restriction enzyme that recognizes the first type IIS restriction enzyme recognition site and cleaves the cDNA upstream of the recognition sequence, thereby generating one or more cDNA fragments,

wherein each of these fragments comprises the oligonucleotide sequence corresponding to an upstream cellular exon directly fused to the marker peptide, the type IIS restriction enzyme recognition site and a portion of the native sequence corresponding to the marker peptide;

iv) adding an adaptor sequence to the end of the unknown oligonucleotide sequence;

v) amplifying by the polymerase chain reaction, the fragments containing the oligonucleotide sequences of the exons fused to the marker peptide with oligonucleotide primers complementary to the adaptor and peptide marker encoding sequences;

vi) cloning and sequencing said amplified fragments; and

vii) comparing the sequence of each oligonucleotide against oligonucleotide sequences in a one or more nucleotide sequence databases thereby identifying one or more fusion proteins present in each subgroup of cells.

10. The method of claim 1, 2, 3, 4, 5 or 6 wherein the identity of the protein fused to the marker peptide is determined using a method of 3'Serial Analysis of Viral Integration (3'SAVI), wherein said method consists of:

i) isolating RNA from each subgroup of cells;

ii) reverse transcribing the mRNA into double stranded cDNA;

iii) subjecting the cDNA to a restriction enzyme that recognizes the second type IIS restriction enzyme recognition site and cleaves the cDNA downstream of the recognition sequence, thereby generating one or more cDNA fragments, wherein each of these fragments comprise a portion of the native sequence corresponding to the marker peptide the type IIS restriction enzyme recognition site and the oligonucleotide sequence corresponding to a downstream cellular exon directly fused to the marker peptide;

iv) adding an adaptor sequence to the end of the unknown oligonucleotide sequence;

v) amplifying by the polymerase chain reaction, the fragments containing the oligonucleotide sequences of the exons fused to the marker peptide with oligonucleotide primers complementary to the adaptor and peptide marker encoding sequences;

vi) cloning and sequencing said amplified fragments; and

vii) comparing the sequence of each oligonucleotide against oligonucleotide sequences in a one or more nucleotide sequence databases thereby identifying one or more fusion proteins present in each subgroup of cells.

11. The method of claims 1, 2, 3, 4, 5, or 6, wherein the peptide marker encoding sequence lacks a translation initiation codon and possesses a translation STOP codon.

12. The method of claims 1, 2, 3, 4, 5 or 6, wherein the peptide marker encoding sequence lacks a translation initiation and STOP codons.

13. The method of claims 3, 4, 5 or 6 wherein said sorting of cells into subgroups of cells based on the levels of expression of the peptide marker is performed by fluorescent activated cell sorting.

14. The method of claims 1, 2, 3, 4, 5 or 6 wherein the oligonucleotide sequence is a fluorescent protein coding oligonucleotide sequence.

15. The method of claim 14, wherein the fluorescent protein encoding oligonucleotide is a green fluorescent protein (GFP) coding sequence.

16. The method of claim 15, wherein the GFP oligonucleotide coding sequence is a humanized renilla GFP (hrGFP) coding sequence.

17. The method of claims 1, 2, 3, 4, 5 or 6 wherein the protein coding sequence is an epitope recognized by fluorescently or enzymatically labeled antibodies.

18. The method of claims 1, 2, 3, 4, 5 or 6 wherein the marker peptide encoded by the polynucleotide requires interaction with another protein in order to generate a fluorescent signal.

19. The method of claims 1, 2, 3, 4, 5 or 6 wherein the polynucleotide construct is introduced into the genome of the cell via a vector.
20. The method of claim 19, wherein the vector is a viral vector.
21. The method of claim 20, wherein the viral vector is selected from the group consisting of a retroviral vector, a lentiviral vector, an adenoviral vector, and an adeno-associated viral vectors.
22. The method of claim 21, wherein a viral vector is selected from the group consisting pGT-fs0, pGT-fs1, pGT-fs2, pHSG, pHG, pHSGEX, pHS2GEX2, pHS2GEX3, pHS2GEX3PA, pHS2AcGEX3, pHS3AcGEX3, pHS3AcGEX4, pHS3AcGEX4PA, pHS3AcGEX4E1bA, pHS3AcGEX4E1bNeo, pGT18, pGT19, pGT22, pGT-GIG, pGT10, pGT11, pGT12, pGT13, pGT14, pGT15, pGT5, pGT5AH and pHS-HA3F.
23. The method of claim 8 wherein following amplification of the one or more extended cDNA fragments, and prior to cloning and sequencing the one or more cDNA fragments, the fragments are ligated together to form a concatenated molecule.
24. The method of claim 9 wherein following amplification of the one or more extended cDNA fragments, and prior to cloning and sequencing the one or more cDNA fragments, the fragments are ligated together to form a concatenated molecule.
25. The method of claim 10 wherein following amplification of the one or more extended cDNA fragments, and prior to cloning and sequencing the one or more cDNA fragments, the fragments are ligated together to form a concatenated molecule.
26. The method of claims 1, 2, 3, 4, 5 or 6 wherein the polynucleotide construct further comprises, downstream of the oligonucleotide encoding a marker peptide an internal ribosome entry site followed by another protein expression marker.
27. The method of claims 1, 2, 3, 4, 5 or 6 wherein the polynucleotide construct further comprises, downstream of the oligonucleotide having a specified sequence, a sequence encoding, upon expression, a selectable marker.
28. A method of screening for small molecule drugs, said method comprising:

- i) generating cells by using methods of claims 1, 2, 3, 4, 5 or 6;
- ii) selecting cells which have integrated the marker peptide into a locus coding a protein for which a small molecule drug is to be identified;
- iii) establishing a monoclonal cell line from cells of step ii); and
- iv) screening the cell line of step iii) against libraries of drug compounds to identify compounds which decrease expression of the marker polypeptide by means of inhibiting expression of the protein to which the marker polypeptide is fused.

29. A method of identifying protein/protein interactions, said method comprising of:

- i) establishing a stable cell line which expresses a fusion protein between a target protein for which the interaction partners are to be identified and a first subunit of a reporter protein system which will be used to assay for the interaction between said target protein and the proteins which interact with said target protein;
- ii) performing a method of claim 1, wherein the marker peptide is a second subunit of said reporter protein system, on said established cell line;
- iii) screening for cells where said reporter system activity has been reconstituted by interaction of said first and said second subunits; and
- iv) identifying a protein interacting with said target protein by any of the methods according to claim 7.

30. The method of claims 1 or 2, wherein the generation of an assayable signal from the peptide marker requires interaction with other protein or protein fusions, said method comprising of:

- i) establishing a stable cell line which expresses a fusion protein between a target protein for which the interaction partners are to be identified and a first subunit of a two-subunit reporter protein system which will be used to

assay for the interaction between said target protein and the proteins which interact with said target protein;

- ii) performing a method of claim 1, wherein the marker peptide is a second subunit of said reporter protein system, on said established cell line;
- iii) screening for cells where said reporter system activity has been reconstituted by interaction of said first and said second subunits; and
- iv) identifying a protein interacting with said target protein by any of the methods according to claim 7.

31. The method of claim 29 or 30, wherein the reporter is  $\beta$ -lactamase.

32. The method of screening for small molecule drugs, wherein the screening of small drug libraries is performed in cells generated according to claim 29 or 30.